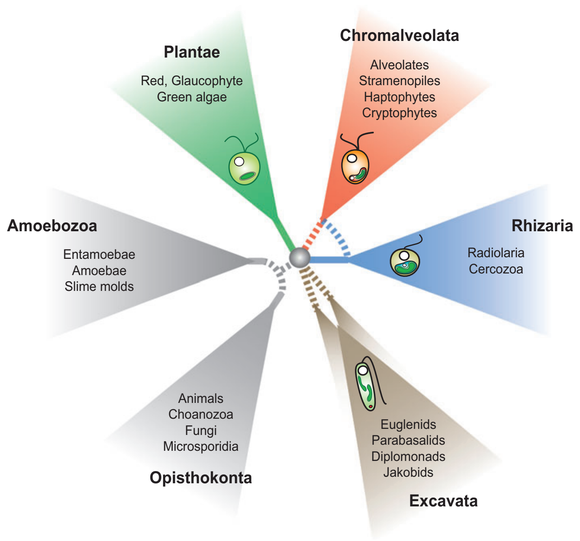
A Phylogenomic Falsification of the Chromalveolate Hypothesis

Denis Baurain†, Henner Brinkmann, Jörn Petersen, Naiara Rodríguez-Ezpeleta, Alexandra Stechmann, Vincent Demoulin, Andrew J. Roger, Gertraud Burger, B. Franz Lang, and Hervé Philippe

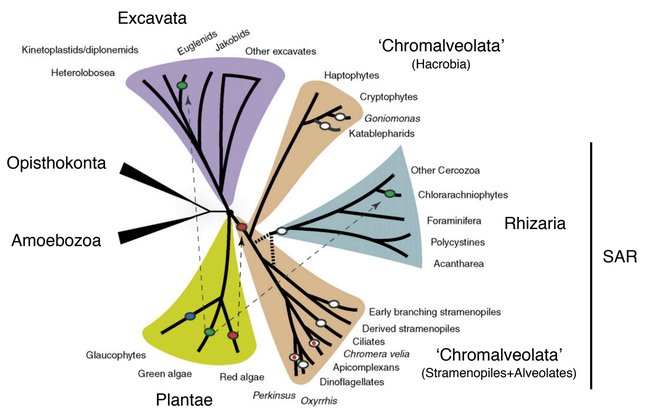
Département de Biochimie, Centre Robert-Cedergren, Université de Montréal, Montréal, Québec, Canada
Département des Sciences de la Vie, Université de Liège, Liège, Belgium
†Present address: Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, Université de Liège, Liège, Belgium

What are 'Chromalveolates'?



CHROMALVEOLATES are a large and diverse putative supergroup of eukaryotes that include a mix of photosynthetic and heterotrophic lineages. According to the Chromalveolate hypothesis (Cavalier-Smith 1999), which regards plastid loss as much more common than plastid gain, these organisms vertically descent from a single chl. c-containing ancestor that acquired its plastid early via a secondary endosymbiosis with a red alga (Reyes-Prieto *et al.* 2007).

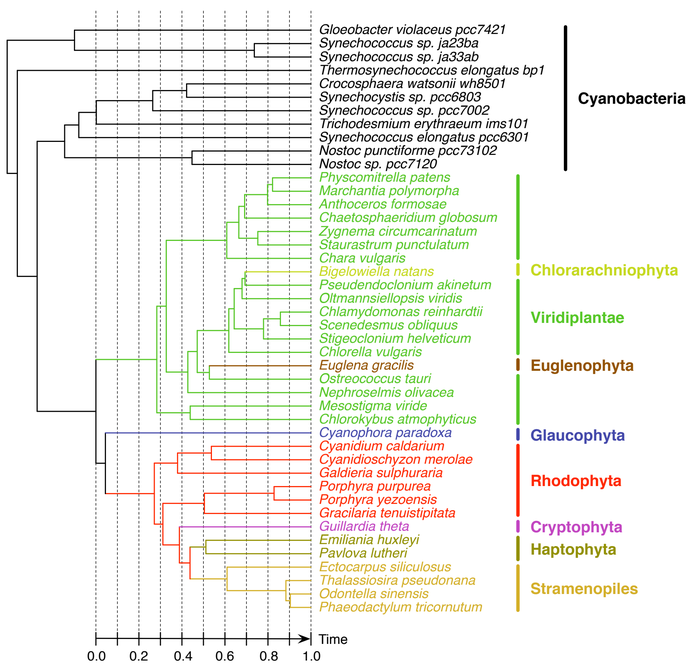
Do 'Chromalveolates' really exist?



RECENT phylogenomic studies have led to dramatically expanded 'Chromalveolates'. To account for the growing collection of heterotrophic lineages apparently related to chl. c-containing algae, the Chromalveolate hypothesis has to postulate ever more plastid losses. Further, it suffers from a recurrent lack of support in most molecular phylogenies, even those based on large data sets. Therefore, alternative evolutionary scenarios have been proposed to explain the origin of 'Chromalveolate' organisms (Lane and Archibald 2008).

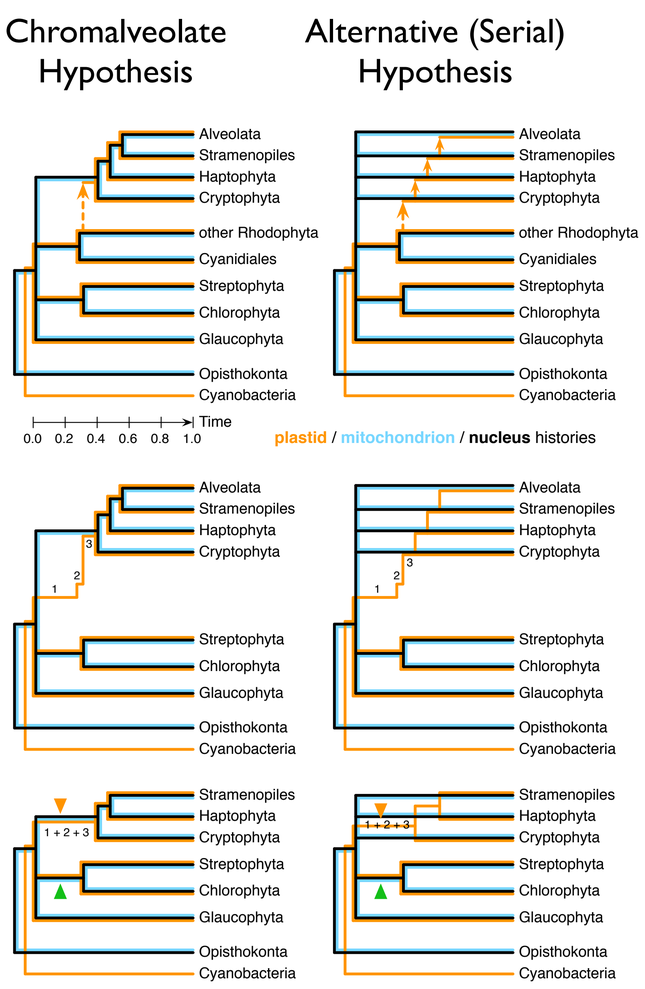
HERE, we present a falsifying experiment that compares the Chromalveolate hypothesis to serial models invoking higher-order eukaryote-eukaryote endosymbioses (EEEs).

How old are chl. c-containing plastids?

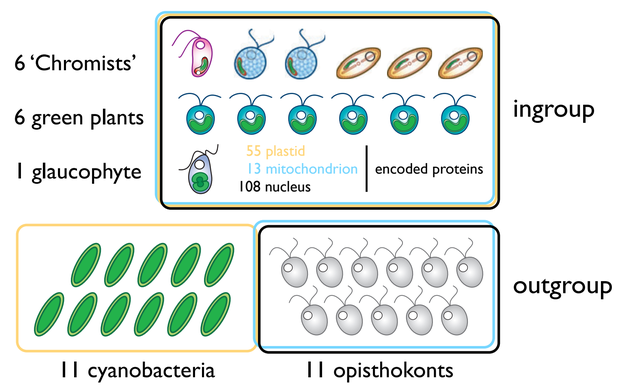


PHYLOGENETIC inference under the CAT model on 55 plastid-encoded proteins (44 OTUs x 10,805 AA) resulted in a tree that was relatively dated using a log-normal autocorrelated model. In this analysis, chl. c-containing plastids appear to have diversified at 38.8% of the time elapsed since the last common ancestor of Plantae and 'Chromalveolates'.

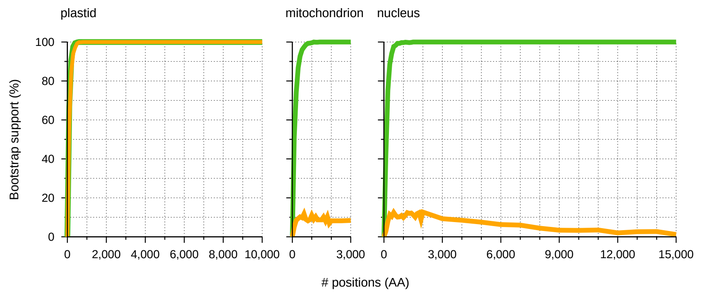
How to test 'Chromalveolates'?



BOTH Chromalveolate and serial hypotheses postulate an initial, single secondary endosymbiosis of a red alga within a eukaryotic host (**dotted arrows**), leading to the emergence of a chl. c–containing founder. The Chromalveolate hypothesis then assumes that this alga gives rise to 'Chromalveolates' by vertical descent (**top left tree**). In contrast, serial hypotheses posit multiple subsequent EEEs (**plain arrows**), which horizontally spread plastids among otherwise unrelated eukaryotes (**top right tree**). As chl. c plastids emerge from within red algae, overall histories of plastid (**orange**), mitochondrial (**blue**), and nuclear (**black**) genomes cannot be superimposed. However, removing red algae 'regularizes' plastid history by creating a single branch out of three smaller ones (**middle trees**). Now, the Chromalveolate hypothesis predicts that the signal for the monophyly of 'Chromists' (discarding alveolates) should be similarly strong across all genomic compartments (**orange arrowhead in bottom left tree**), while it should be strong only with plastid genomes in serial hypotheses (**orange arrowhead in bottom right tree**). To validate our approach, green plants were used as a test case, since the signal for their monophyly is expected to be similarly strong, regardless of the compartment or the hypothesis (**green arrowheads**).

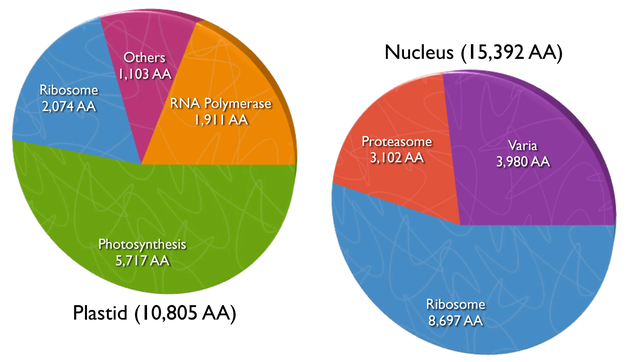


USING the same sampling of ingroup species, we assembled one concatenated protein data set per compartment. The strength of the phylogenetic signal was then estimated through a Variable Length Bootstrap (VLB) strategy.



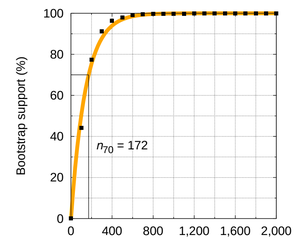
WHATEVER the compartment considered, the monophyly of green plants is easily recovered, whereas the monophyly of 'Chromists' is only recovered with plastid genomes. This observation falsifies the Chromalveolate hypothesis.

How to check our assumptions?



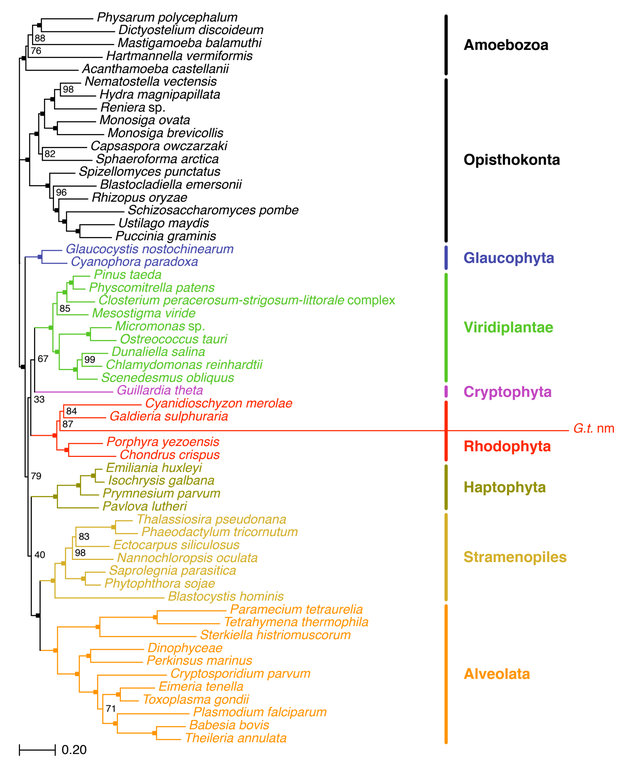
SINCE heterogeneous rates may affect inference, we decided to subdivide each of our plastid and nuclear data sets into smaller data sets according to functional class.

THEN, to allow easy comparison, we defined n_{70} as the number of positions required to reach a VLB support ≥70% for the monophyly of the group put to test (by fitting a monomolecular model).



	# Pos	f(Sub)	Trees	VLBs: Greens	VLBs: 'Chromists'	n_{70} Values: Gre	n_{70} Values: 'Chr'
Plastid							
Polymerase	1,911	4.79				365	1,802
Photosynthesis	5,717	1.63				101	95
Ribosome	2,074	3.56				208	172
Mitochondrion	3,106	3.51				176	n.c.
Nucleus							
Proteasome	3,102	2.34				501	n.c.
Ribosome	8,697	2.72				130	n.c.
Varia	3,980	1.99				305	n.c.

ALTHOUGH plastid genomes display extreme rate variations, the monophyly of 'Chromists' is recovered with n_{70} values that remain small relative to the number of positions available. In contrast, VLB support is so low in mitochondrial and nuclear compartments that it is not even possible to fit the model. Thus, rate heterogeneity does not impair our test.



FINALLY, using an extended nuclear data set (57 OTUs x 15,392 AA), we tested the affinity of the fast-evolving and compositionally biased nucleomorph of *Guillardia theta*. Inference under the CAT model yielded 100% bootstrap support for the grouping of the nucleomorph with red algae, while providing no support for the monophyly of 'Chromalveolates'. This helps us to exclude phylogenetic artifacts and to conclude that the Chromalveolate hypothesis is falsified.

Want to know more?

BAURAIN D. *et al.* (2010) Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol. Biol. Evol.* **27**:1698-1709.

SMBE2010, Annual Meeting of the Society for Molecular Biology and Evolution, Lyon, France, July 4–8, 2010